AD NUMBER

AD455410

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; OCT 1964. Other requests shall be referred to Personnel Research Lab., Aerospace Medical Div., AFSC, Lackland AFB, TX.

AUTHORITY

AFMPC ltr, 17 May 1972

PRL-TR-64-29

October 1964

Officer Promotion Procedures

III. Increasing the Reliability of Promotion Board Evaluations

by
Ernest C. Tupes
Ray W. Alvord, Maj USAF
Lonnie D. Valentine, Jr.

PERSONNEL RESEARCH LABORATORY
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
Lackland Air Force Base, Texas

AFPS 8A

NOTICE

All distribution of this report is controlled. Qualified Defense Documentation Center users will request through Headquarters USAF, Promotion Board Secretariat (AFPMPJA).

Defense Documentation Center release to the Office of Technical Services is not authorized (see Foreword).

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

VARITYPIST: ESTHER BARLOW

INCREASING THE RELIABILITY OF PROMOTION BOARD EVALUATIONS

I. INTRODUCTION

In August 1962 a series of studies of officer promotion actions were initiated to analyze officer promotions from the standpoint of reliability and stability; to devise and carry out analyses and experimental studies leading to possible increases in the efficiency of the officer promotion system; and to propose for tryout those changes which the analyses indicate would feasibly result in increased efficiency.

The present report is the third in a series describing the analyses and studies carried out. In this report ways in which the reliability of the officer promotion system might be increased are discussed.

Analyses of a number of FY 1962 officer Promotion Boards[1] have indicated that Promotion Scores and the resulting recommendations concerning promotion or nonpromotion of individual officers are quite reliable and compare favorably in reliability with other types of ratings and judgments. The analyses have indicated, however, that a small amount of unreliability is present in the Promotion Scores and promotion recommendations. This unreliability, although low in comparison with many other types of judgments, is sufficiently high so that from 8 to 21 percent (depending upon the board studied) of the recommendations (to promote or not to promote) would be reversed were the records of the same group of eligible officers to be reevaluated by another Promotion Board. That is to say, promotion or nonpromotion of a percentage of eligible officers, in any given cycle, is based only partly on their past performance as Air Force officers and partly on whether their records are assigned to one panel or board for evaluation rather than to some other panel or board. Unreliability cannot ever be entirely eliminated and, as indicated above, has as little or less influence in the Promotion Board system than in other judgmental and evaluation situations. However, the reliability of the present promotion decisions can be increased and the influence of chance factors reduced to a minimum with certain changes in the evaluation system.

The purpose of this report is to note certain factors which are related to reliability of judgments and to discuss ways in which these factors could be brought to bear to increase the reliability of the present Promotion Board system.

II. NUMBER OF RATERS

Many studies have shown that evaluations based on the average or sum of evaluation scores from a number of raters are more reliable than evaluations from a single rater; and, further, that the reliability increases as the number of raters is increased. Presently, most Promotion Board evaluations are accomplished by three raters. In Table I are presented estimates of the reliability which would be expected were more (or fewer) than three raters to evaluate each selection folder. These estimates are based on two FY 62 Promotion Boards and are expressed in terms of the percentages of eligibles for whom the promotion recommendation made by one board would be reversed were the folders to be reevaluated

---

[1] L. D. Valentine, Jr., & E. C. Tupes. *Officer promotion procedures: I. An analysis of officer promotion actions.* PRL-TR-64-27. Personnel Research Laboratory, Aerospace Medical Division, October 1964.

Table I. **Reliability of Promotion Recommendations as a Function of Number of Raters**

NUMBER OF RATERS	PERCENTAGE OF REVERSALS[a]	
	BOARD A[b]	BOARD B[c]
1	15.2	31.4
2	9.7	24.6
3	7.9	20.6
4	6.3	18.2
6	4.9	14.6
9	4.2	12.6
12	2.8	10.9
24	1.6	1.9

[a] The percentage of eligibles for whom the promotion recommendation would differ were each folder to be reevaluated by a second board with the same number of raters per folder.

[b] Board A is a FY 62 temporary board which promoted about 78% of the eligibles, with an estimated reliability of $r = .90$.

[c] Board B is a FY 62 temporary board which promoted about 44% of a prescreened group of eligibles, with an estimated reliability of $r = .77$.

by a second board of equivalent size. It can be seen from Table I that the unreliability of the promotion recommendations could be sharply reduced by increasing the number of raters who evaluated each selection folder. For example, looking at Board B, it can be seen that the 20.6 percent of reversals in recommendation expected with the present three-rater panel would be reduced to 18.2 percent with four raters per folder and to 14.6 percent with six rarers per folder. Were each folder to be evaluated by only one rater, the reversals would be expected to increase to 31.4 percent.

To increase the panel size from three to six or nine or more raters per folder would probably involve an unacceptable increase in the expense (in terms of personnel and travel costs) of the system. However, the number of raters could be increased by use of a dual-board system, wherein the eligibles in any command could be evaluated by a board at command headquarters and again by a Headquarters USAF board. Use of the same scoring and evaluation procedures by the command and the Headquarters USAF boards would permit the evaluation scores from both boards to be averaged together to obtain a final Promotion Score.[2] Since this final score would be based upon evaluations from six raters an increase in reliability would result. The system could be extended to include base level boards if desired so that the final scores would be the sum of nine ratings and consequently more reliable. Quality control procedures could easily be incorporated into the system if it were believed desirable to rule out the possibility of command or base differences in evaluation and rating standards.

A dual-board system is presently in effect for promotions to temporary colonel so that evaluations from six raters are already available for this group. However, the first board acts only as a Nominating Board and the final Promotion Scores are based only on the Promotion Board ratings; thus the reliability is essentially only that of a three-rater system. If the Nominating Board scores

[2] This suggestion is not for a return to the system wherein eligibles were screened by the command and the better qualified nominated for Headquarters USAF consideration. That system had the reliability of a three-rater system whereas the suggested system would have the reliability of a six-rater system.

for nominees were added to the Promotion Board scores final promotion recommendations would have the reliability of a six-rater system. In view of the present availability of the ratings and the slightly lower reliability found for the temporary colonel promotions, this proposed procedure might be considered, regardless of whether it is deemed feasible to establish any sort of dual-board system for other grades.

III. RELIABILITY OF HIGH AND LOW PROMOTION SCORES

An assumption of reliability estimates is that the reliability of a score does not vary with its level—that is, high scores, low scores, and average scores are equally reliable. However, the reliability of promotion recommendations does vary with the level of the Promotion Score. Eligibles with scores well above any cutting score would also have scores above the cutting score if reevaluated by a second board. Eligibles with scores considerably below the cutting score would also have scores below the cutting score if reevaluated by a second board. The closer any score to the cutting score, the greater the probability that a second board would assign a score on the other side of the cutting score. This point is illustrated by the data in Table II. It can be seen that the reliability of the recommendations is estimated to be 100 percent for eligibles receiving scores of 26 and above or scores of 17 and below. However, the reliability of recommendations for eligibles receiving scores of 22 is estimated to be about 67 percent and about 60 percent for eligibles receiving scores of 21.

Table II. **Reliability of Promotion Recommendations as a Function of Distance of Promotion Score from Cutoff Point[a]**

PROMOTION SCORE	PERCENTAGE DISTRIBUTION	% ABOVE CUTOFF[b]	% BELOW CUTOFF[c]	PERCENTAGE DISTRIBUTION OF REVERSALS[d]	PERCENTAGE OF ALL ELIGIBLES REVERSED
30	0.9	100.0	0	0	0
29	2.0	100.0	0	0	0
28	4.6	100.0	0	0	0
27	8.3	100.0	0	0	0
26	12.1	100.0	0	0	0
25	14.7	99.6	0.4	0.7	0.1
24	14.6	97.2	2.8	6.6	0.5
23	12.6	88.8	11.2	11.6	1.0
22	9.3	66.7	33.3	31.1	2.5
21	7.4	40.3	59.7	36.7	3.1
20	5.2	16.1	83.9	9.4	0.8
19	3.8	7.1	92.9	3.7	0.3
18	2.4	0.6	99.4	0.2	0.1
17	1.4	0	100.0	0	0
16	0.7	0	100.0	0	0
15	0.3	0	100.0	0	0
	100.0			100.0	8.4

[a] Board A of Table L. In practice, cutoff points are established for each panel so that equal percentages of eligibles considered by each panel will be recommended for promotion. In this table, scores from all panels were put into a common distribution and that cutoff point (22) selected which would have resulted in 78 percent of the eligibles being recommended for promotion.

[b] Percent of eligibles with scores at each level who would be assigned scores of 22 or above by a second board.

[c] Percent whose scores from the second board would be below 22.

[d] A reversal is an eligible who was above the cut-off but who would have been evaluated below the cut-off by a second board, and vice versa.

This factor can be utilized to increase the reliability of the promotion system without any increase in the overall size of any board. The recommended procedure would be to have the selection folders each evaluated first by only one rater, using a scale ranging from scores of 15 to 30. The distribution of one-rater Promotion Scores would be obtained and a tentative cutting score established. Eligibles whose scores were several points above or below the cutting score could be definitely recommended for promotion or for nonpromotion, since evaluations by additional raters would be very unlikely to result in reversals. Eligibles with scores near the cutting score would be evaluated by full panels, which could consist of five or six members at no increase in overall board size. Thus the reliability of scores near the cutting score would be increased. The result would be an increase in the overall reliability of promotion recommendations with no increase in the size of the board.

IV. USE OF A COMPOSITE SCORE BASED ON PAST PERFORMANCE

Other analyses of FY 62 Promotion Boards[3] have demonstrated that both the Mean OER (average of all OERs received) and a Prediction Score (based on the Mean OERs and other past performance variables) are highly related to Promotion Scores and to promotion recommendations. This relationship is illustrated in Table III for one FY 62 board. This table indicates that Predicted Scores of 26 and higher or 18 and lower were 100-percent accurate in identifying eligibles who were recommended for promotion or for nonpromotion. As the composite scores approach the middle of the range, the accuracy progressively decreases.[4]

Either the Mean OER or a Predicted Score could be incorporated into the promotion system with the result that the reliability of the promotion recommendations could be increased and the number of board members reduced. There are several ways in which this could be done.

One approach is a Zoning Method. Based on either Mean OERs or Predicted Scores, zones of eligibles would be established. Eligibles whose scores were above a specified level would be placed in the Upper Zone. Eligibles whose scores were below a specified level would be placed in the Lower Zone. All other eligibles would be placed in the Gray Zone. Upper and Lower Zone eligibles would be evaluated by one rater. If the rater agreed with the Zoning (Upper Zones to be promoted, Lower Zones not to be promoted), no further evaluation would be required. Eligibles on whom the rater disagreed with the Zoning would be placed in the Gray Zone. All Gray Zone eligibles would be evaluated by a panel of three (or more for greater reliability) raters. A modification of this Zoning Method has been used with two Captains Boards and found to result in a significant decrease in the number of evaluators required overall.[5]

A second approach is a Linear Ordering Method. A Predicted Score would be computed for each eligible. Each eligible would then be evaluated by one rater who would indicate whether he believed the Predicted Score to be approximately correct or whether it was too high or too low. Predicted Scores for the former group would be allowed to stand as the final Promotion Score. Eligibles with whose Predicted Score the rater disagreed would be evaluated by a panel who would assign the final Promotion Scores.

---

[3] R. W. Alvord & E. C. Tupes. *Officer promotion procedures: II. Feasibility of computer applications in the promotion of USAF officers.* PRL-TR-64-28. Personnel Research Laboratory, Aerospace Medical Division, October 1964.

[4] Of interest is the fact that if a composite score of 23 (selected because it comes closest to "promoting" the same percentage as were actually promoted) is used as a cutting score, the overall accuracy is about 85 percent. The overall inaccuracy (15%) is about that expected (see Table I) from evaluation scores based on one rater. Thus it might be inferred that the composite is as accurate or "reliable" as a single rater.

[5] Promotion Board Secretariat, Hq USAF.